

APPLICATION FOR UNITED STATES LETTERS PATENT FOR

MULTIBYTE LOCALE FOR SINGLE BYTE LANGUAGES

INVENTORS:

Hiroshi Watanabe
William A. Shannon

PREPARED BY:

David L. Stewart, Esq.
LOWE, PRICE, LeBLANC & BECKER
99 Canal Center Plaza, Ste. 300
Alexandria, VA 22314
(703) 684-1111

2860-005

P1222

BACKGROUND OF THE INVENTION

Field of the Invention

5 The invention relates to the development and testing of software to be deployed internationally and more particularly to the development and testing of software for languages requiring a multibyte representation for characters.

Description of Related Art

10 As computer hardware and software vendors expand their markets to Europe and the Far East, they are required to modify the related operating system and applications software to accommodate the language, customs and culture of the individual target markets. Many of these companies now achieve half of their total revenues from such markets.

15 The development and deployment of international software is discussed in a book entitled Developing and Localizing International Software, by Tom Madell et al.

published in 1994 by Prentis-Hall, Inc., of Englewood Cliffs, New Jersey.

5 Much software is designed and programmed taking into consideration only the needs of users in the particular country where it is developed. As a result, other international users of the software are forced to struggle with the language of development and its corresponding cultural representations in order to use the software or, software designers or engineers must redesign and recompile the software to create a new version for each unique language and local environment that uses it.

10 An approach that enhances software for worldwide distribution uses internationalization and localization. Internationalization is sometimes referred to as I18N, for the eighteen letters between the I and the N in "internationalization," and is a process of configuring a program to make localization easy. Similarly, localization is sometimes referred to as L10N and generally involves more than merely converting languages of messages and displays.

20 Developers for a worldwide audience must enable computer systems to read and write in the user's native language, that is, to understand and display characters and symbols that may be far different from the characters set of a single byte language such as American English.

Further, the computer systems must be enabled to have the capability to process the characters and text according to the rules of the user's language. Many languages have characters in excess of the twenty-six characters of the English language set. Software to be used internationally must provide flexibility to modify output conventions to comply with customary local requirements for representations of currency, numeric data or time. Such software should also provide the ability to allow for the translation of interfaces, messages and prompts without necessitating many different language versions of the underlying software.

I18N then, is the process of building in the capabilities which facilitate adaptation to different countries and locales during the development or modification process.

While I18N is usually a process performed during development of the code, localization, or L10N, is most often carried out subsequent to development, often in the foreign location where the software will be utilized. L10N is a process of actually adapting the potentially useful internationalized software to meet the needs of one or more users in a particular geographical area. It includes not only the translation of messages but also the selection or creation of appropriate language tables containing the relevant local data for use on a given

system. Localization activities are usually performed by the software manufacturer or its representative in a particular locale.

5 Much software has been developed for use with terminals that generate 128 ASCII characters which can be represented by 7 bits of a single byte. Some other languages require 256 characters which require all 8 bits of a single byte. Such languages can be described as single byte languages. Some code sets for other languages, such as some Asian languages, contain 10 thousands of characters and require more than a single byte. These can be described as multibyte languages.

Thus, localization must be possible for multibyte languages. This resulted in the development of worldwide portability interfaces (WPI) as defined by X/OPEN in 15 standard XPG4. To internationalize a program, developers no longer deal with character data in a language-sensitive way; nor do they need knowledge of any foreign language. They do not even need to be aware of the ways 20 different languages and code sets can vary. This makes the programming effort easier and provides for consistent treatment across languages.

A central concept of internationalization and localization is that a developer should be able to create 25 a single internationalized application, which is capable

of being extended into any number of localized programs without the need for redesign or recompilation.

To enable this, an internationalization model includes three parts. Namely, a language independent program, message catalogs and language tables. **Figure 1** illustrates a model of internationalized software. A language independent program 100 achieves language independence by programmatic calls to a message catalog 110 and to language table 120. Rather than hard-coding messages such as prompts and error messages within the program itself, such messages are stored in external message catalogs with a different version of those catalogs for each supported language. Language tables contain all language-specific processing information and conventions unique to a particular locale, such as how characters are sorted and how output (such as numbers, times and dates) is formatted. At run time, generally in a development environment the program selects or "binds" a specific language table according to settings controlled by the user, the application developer, or system administrator. Thus, the same basic program 100 can be executed in different language "locales" by simply binding the appropriate message catalog and language table to the program at run time.

The term "locale" will be utilized to refer to the language table component of an internationalized application.

5 Use of the internationalization model of **Figure 1** provides several advantages. First, software does not need to be recoded in different versions for different languages in order to localize it. As a result, only one version needs to be updated and maintained as well as manufactured, stocked and shipped.

10 Because all language-dependent information is kept external to the program source, neither programmers nor translators need to modify the program source code in order to localize program language behavior. As a result, the possibility of introducing bugs into the main
15 or core program during localization of the software is eliminated. Instead of having to work with a huge amount of source code, translators can work from a discrete message file containing only the text to translate. This frees them from the need to have programming skills.

20 Since software can be localized more easily and inexpensively, such software should become more readily available and, as a result, fewer end users will be forced to deal with difficult foreign language representations.

25 By using external language tables, the structural and processing rules of each language are consolidated

into one physical location which can be modified to meet even more specific local requirements. Once installed, such language tables can support other internationalized programs on the system.

5 Different cultures and countries have different rules for punctuation, word order, sorting, the order of items and addresses, currency measures and conversions, number formats and other local idiosyncracies. Many native languages and customs have different meanings for
10 certain symbols used as computer icons as well as colors which may be used to indicate some special meaning.

 Localization of a computer product from one locale to another to accommodate such differences more specifically involves:

- 15 1. Translation of software documentation into the new language;
2. Translation of the textual messages embedded in the software into the new language;
- 20 3. Incorporation of additional software facilities to make input and output of the new language and perhaps new characters possible;
4. Adapting the software to
25 accommodate the customs and conventions of the new locale; and

5. Testing and assurance that the modified product works as intended in the new locale.

This process of localization is very labor intensive and requires people who know the native language of the new country as well as the basics of computer program architecture and construction.

Figure 2 illustrates an internationalized program which has been localized into a plurality of languages. Program 200 has been internationalized, that is, written with the appropriate hooks so that a particular message catalog 210 and a corresponding language table 220 can either be bound to the program 200 at run time or selected by virtue of a software switch. Note that languages such as Korean (220C) and Japanese (220I) require a character set which cannot be represented with a single byte of data. A language table which requires that each character be represented by more than one byte is called a multibyte locale. Typically, a two byte representation or a wide character is utilized for each character in a multi-byte locale. Languages such as French (220A) and German (220B) are single byte languages which have a character set which can be represented in a single 8 bit byte. American English is also a single byte language which can be represented in 7 bits of an 8 bit byte and such a representation is referred to as USASCII.

Figure 3 illustrates development of an internationalized computer program in the U.S.A. As the internationalized computer program 300 is developed, a catalog of English messages is concurrently developed (310). If a language table or locale 320 which reflects the customs of the locale of development, namely, the United States, has been developed, there is no need to repeat the development. Only USASCII need be supported and the time representations, dates, currency formats, sort order and the like are those in use in the U.S.A. A set of software development tools 340, permit the development, debugging and compiling of the I18N software 300 and the creation of the message catalog 310 and the USASCII English locale 320. In their simplest form, the development tools would include text editor for creating source code, message catalogs and locales , and a compiler.

Figure 4 illustrates life cycle development and testing of an internationalized computer program developed in the United States. The development of an I18N program, English message catalog in U.S. ASCII English locale (400) proceeds concurrently as shown in **Figure 3**. At various stages during development, the program is tested in its native locale (410). If bugs are found, they are corrected in the development phase 400. The local testing step 410 encompasses both

informal testing done by the developer and formal testing as the result of a release. There may be several iterations of development testing and changes (400, 410) until the development is considered sufficiently stable to be passed to a localization team for localization to a non-U.S. locale (420). Development of the localized version 420 and testing of the localized version 430 may result in the discovery of software bugs unique to the localization process. These will be corrected by the localization team (420). However, such testing may also reveal problems with the development of the I18N core program itself. Such problems must, therefore, be referred back to the developers for correction (400).

The Problems

The development process outlined above has several deficiencies. First, many software bugs result when attempting to localize an internationalized software program to a multi-byte locale. These are not identified until testing of the localized version and as a result, feedback to developers occurs long after the introduction of the error and after the time and effort has been expended to release the software to a localization team. Such late identification of errors greatly increases the cost of correction of the software.

Further, enhancements of functionality and incorporation of engineering change orders into the

software cannot be tested in the multi-byte version until
release of the software to the localization team. This,
too, increases the cost of development and maintenance of
the software. There is thus a need for improving the
5 development and testing process of internationalized
software.

SUMMARY OF THE INVENTION

10 The present invention provides a system, process and
computer program product which has the advantage of
permitting testing of much of the multi-byte
functionality during the development process. This is
accomplished by providing a multi-byte locale for a
single byte language which would act for testing purposes
just like a multi-byte locale for a multibyte language
15 but in which the content was in the single byte language.
Thus, an I18N program developer using a single byte
language such as English could develop the I18N core
program, an English message catalog and utilize a USASCII
English locale and then immediately test the development
20 work in a multi-byte English locale which should produce
the identical results found in the USASCII English locale
represented, however in multi-byte or wide characters.
Thus, many of the errors associated with the localization
process can be detected early and corrected by the
25 developer prior to release to the localization teams.
Further, enhancements to the software and incorporation

of engineering change orders can be developed and tested in both the USASCII locale and in a multi-byte English locale, thus identifying and correcting at an early stage those software bugs at much less cost.

5 The invention is directed to a method of testing internationalized software, by binding internationalized software to be tested to a multibyte locale created for a single byte language. In it's simplest form, this is done by creating a mapping between multibyte binary words and characters of said single byte language; and providing for conversion of representations of characters of said single byte language into corresponding multibyte binary words specified by said mapping. Another form includes providing a sort function which sorts multibyte binary words in a sort order customary for said single byte language, but which differs from a binary sort order; as a result of which, failure to invoke said sort function of the multibyte locale will result in a different sort order from said sort order customary for said single byte language. Yet another for includes defining a date representation for a particular locale; and providing a date function which converts an internationalized date representation to said date representation for a particular locale as a result of which failure to invoke said date function of the multibyte locale will result in a different date

representation from said date representation customary for said locale. It is desirable to provide for display of multibyte binary words so as to create a visual distinction between characters represented in said multibyte binary words and characters represented in ASCII. The visual distinction can relate to one of font, color or spacing.

The invention is also directed to apparatus for testing internationalized software, including a computer; and a computer program stored on said computer for use with said internationalized software, said program including a multibyte locale created for a single byte language.

The invention is also directed to a computer system for developing software having at least one computer; one or more message sources, each containing one or more program messages in a single byte language; and one or more language tables containing specific processing information and conventions for a particular locale, at least one of which is a multibyte locale created for a single byte language.

The invention is also directed to a computer system for developing and testing an internationalized computer program written in a single byte language including a network, one or more computers connected to the network;

a source of messages to be used by said internationalized computer program when running on said one or more computers; and one or more locales, each containing specific processing information and conventions for a particular location, at least one of which is a multibyte locale created for said single byte language, for binding to said internationalized computer program during program development and testing, by which an internationalized computer program under development can be tested

10 The invention is also directed to a product for implementing a multibyte locale on a computer readable memory medium by including a data structure stored on said memory medium, utilized for controlling the multibyte locale, said data structure having a mapping of characters of a single byte language to corresponding multibyte binary words, a mapping of elements of a date representation utilized with internationalized software to elements of a date representation of a particular locale or a representation of sort order utilized in a particular locale.

20 The invention is also directed to a computer program product for implementing a multibyte locale including a computer readable memory medium; and one or more language tables containing specific processing information and conventions for a particular locale, at least one of

which is a multibyte locale created for a single byte language.

The invention is also directed to a computer program product for implementing a multibyte locale in a single
5 byte language comprising a computer readable memory medium; and a computer program including a routine for conversion of representations of characters of said single byte language into corresponding multibyte binary words, a routine providing a sort function which sorts
10 multibyte binary words in a sort order customary for said single byte language, but which differs from a binary sort order; and a routine for providing a date function which converts an internationalized date representation to said date representation for a particular locale.

15 The invention is also directed to a network with improved capabilities for testing internationalized software, which has a plurality of computers connected to the network and at least one of said computers configured to bind an internationalized program written in a single
20 byte language to a multibyte locale created for said single byte language.

The invention is also directed to a method of testing internationalized software written in a single
25 byte language using a network by downloading, over said network, a multibyte locale implemented in said single

byte language; and binding said multibyte locale to said internationalized software for testing.

5 The invention is also directed to a method of facilitating testing of internationalized software written in a single byte language at a remote location using a network by sending, over said network, a multibyte locale created for said single byte language to a computer at said remote location, as a result of which said computer at said remote location can bind the multibyte locale created for said single byte language to said internationalized software for testing.

10 Still other objects and advantages of the present invention will become readily apparent to those skilled in the art from the following detailed description, wherein only the preferred embodiment of the invention is shown and described, simply by way of illustration of the best mode contemplated of carrying out the invention. As will be realized, the invention is capable of other and different embodiments, and its several details are capable of modifications in various obvious respects, all without departing from the invention. Accordingly, the drawing and description are to be regarded as illustrative in nature, and not as restrictive.

BRIEF DESCRIPTION OF DRAWINGS

The objects, features and advantages of the system of the present invention will be apparent from the following description in which:

5 **Figure 1** illustrates a model of internationalized software.

Figure 2 illustrates an internationalized program which has been localized into a plurality of languages.

10 **Figure 3** illustrates development of an internationalized computer program in the United States of America.

Figure 4 illustrates life cycle development and testing of an internationalized computer program developed in the United States.

15 **Figure 5** illustrates an exemplary development environment containing a multi-byte locale for a single byte language in accordance with the invention.

Figure 6 illustrates an improved test method in accordance with the invention.

20 **Figure 7A** illustrates a computer of a type suitable for carrying out the invention.

Figure 7B illustrates a block diagram of the computer of **Figure 7A**.

25 **Figure 7C** illustrates a floppy disk containing a program usable with the computer of **Figure 7A**.

NOTATIONS AND NOMENCLATURE

The detailed descriptions which follow may be presented in terms of program procedures executed on a computer or network of computers. These procedural descriptions and representations are the means used by those skilled in the art to most effectively convey the substance of their work to others skilled in the art.

A procedure is here, and generally, conceived to be a self-consistent sequence of steps leading to a desired result. These steps are those requiring physical manipulations of physical quantities. Usually, though not necessarily, these quantities take the form of electrical or magnetic signals capable of being stored, transferred, combined, compared, and otherwise manipulated. It proves convenient at times, principally for reasons of common usage, to refer to these signals as bits, values, elements, symbols, characters, terms, numbers, or the like. It should be noted, however, that all of these and similar terms are to be associated with the appropriate physical quantities and are merely convenient labels applied to these quantities.

Further, the manipulations performed are often referred to in terms, such as adding or comparing, which are commonly associated with mental operations performed by a human operator. No such capability of a human operator is necessary, or desirable in most cases, in any

of the operations described herein which form part of the present invention; the operations are machine operations. Useful machines for performing the operation of the present invention include general purpose digital computers or similar devices.

The present invention also relates to apparatus for performing these operations. This apparatus may be specially constructed for the required purpose or it may comprise a general purpose computer as selectively activated or reconfigured by a computer program stored in the computer. The procedures presented herein are not inherently related to a particular computer or other apparatus. Various general purpose machines may be used with programs written in accordance with the teachings herein, or it may prove more convenient to construct more specialized apparatus to perform the required method steps. The required structure for a variety of these machines will appear from the description given.

DESCRIPTION OF THE PREFERRED EMBODIMENT

Figure 5 illustrates a development environment containing a multi-byte locale for a single byte language in accordance with the invention. Blocks 500, 510, 520 and 540 correspond to blocks 300, 310, 320 and 340 of **Figure 3**. The inventors have recognized the fact that the capability of developing software for international

deployment can be greatly enhanced by inclusion in the development environment of a multi-byte locale 530 for a single byte language. This example permits the I18N core program to be run not only in a USASCII English locale (520) but also in a multi-byte English locale, 530. This permits immediate testing of much of the multi-byte functionality during the development of the program on a selective basis. This enables a developer to identify problems very quickly and to correct them immediately, rather than waiting for completion of localization development and testing.

Figure 6 illustrates a revised testing process used when a multi-byte locale for a single byte language is available for testing.

Figure 6 assumes that a standard USASCII English locale has been previously developed. The requirements for locales have been set forth more in detail above. **Figure 6** also assumes that a multi-byte English locale has been previously created. This could be as simple as modifying a U.S. ASCII English locale to accommodate multi-byte or wide characters. However, certain enhancements are preferred in the implementation of the multibyte locale for a single byte language. When creating a multibyte locale for a single byte language, it is desirable to build in features which permit ready identification of errors when invoking operations within

the multibyte locale. One such technique involves creating a visual distinction between a display of regular ASCII characters and a display of a multibyte character. Such a visual distinction could occur in either font, color, or character spacing. Thus, when an ASCII character is displayed without properly invoking the multibyte functions, it would be displayed in normal ASCII mode without the visual distinctions which would be expected had the multibyte locale function been properly invoked.

Similarly, it is desirable to have a special sorting routine for the multibyte locale which sorts in an order different from the normal binary sort order so that if sorting occurs by some medium other than a sort routine of the multibyte locale, the resulting sort order would be noticeably improper. For example, if sorting occurred by a binary sort routine, the order could be reversed from the normal sort routine provided by the multibyte locale, to indicate that the sort mechanism of the multibyte locale had not been properly invoked. If the sort routine of the multibyte locale had been properly invoked, then sorting would occur in the expected order.

The same type of approach can be taken with dates so that when the date routine of the multibyte locale is properly invoked, the date presentation is correct for

the locale, whereas if it is not, the date presentation would be recognizably different.

As shown in **Figure 6**, during development of an I18N program and its associated English message catalog, at various times the product developed will be tested using the USASCII English locale (610A). If bugs are found, the developer can correct them immediately (600) and then retest (610A) to see if the software operates satisfactorily. If it does, the developer then may wish to test using a multi-byte English locale (610B) to identify problems with the I18N core program which might not handle the multi-byte representations properly. Typically, many such problems are found in the display and printing routines. Using the multi-byte English locale, the developer can immediately test for these problems and correct them before the software is released to the localization team (620) and before localized version testing (630).

In this manner, a great many more software bugs are identified early in the development process where they can be corrected with considerably less cost and without the extra work associated with release to a localization team. The existence of a multi-byte locale for a single byte language for testing is particularly advantageous when localization in a number of countries proceeds concurrently. In those countries requiring multi-byte

representations, each localization team would discover the same errors during testing and the costs of identifying those errors would multiply in accordance with the number of teams working concurrently. By allowing the developer to test in a multi-byte locale directly, the developer can test the multi-byte capabilities and still have the test results presented in the single byte language used for development.

Figure 7A illustrates a computer of a type suitable for carrying out the invention. Viewed externally in **Figure 7A**, a computer system has a central processing unit 700 having disk drives 710A and 710B. Disk drive indications 710A and 710B are merely symbolic of a number of disk drives which might be accommodated by the computer system. Typically, these would include a floppy disk drive such as 710A, a hard disk drive (not shown externally) and a CD ROM drive indicated by slot 710B. The number and type of drives varies, typically, with different computer configurations. The computer has the display 720 upon which information is displayed. A keyboard 730 and a mouse 740 are typically also available as input devices over interface 745. Preferably, the computer illustrated in **Figure 7A** is a SPARC workstation from Sun Microsystems, Inc.

Figure 7B illustrates a block diagram of the internal hardware of the computer of **Figure 7A**. A bus

750 serves as the main information highway interconnecting the other components of the computer. CPU 755 is the central processing unit of the system, performing calculations and logic operations required to execute a program. Read only memory (760) and random access memory (765) constitute the main memory of the computer. Disk controller 770 interfaces one or more disk drives to the system bus 750. These disk drives may be floppy disk drives, such as 773, internal or external hard drives, such as 772, or CD ROM or DVD (Digital Video Disks) drives such as 771. A display interface 775 interfaces display 720 and permits information from the bus to be displayed on the display. Communications with external devices can occur over communications port 785.

Figure 7C illustrates an exemplary memory medium which can be used with drives such as 773 in Figure 7B or 710A in Figure 7A. Typically, memory media such as a floppy disk, or a CD ROM, or a Digital Video Disk will contain, *inter alia*, a multi-byte locale for a single byte language and the program information for controlling the computer to enable the computer to perform its testing and development functions in accordance with the invention.

Thus, using the techniques of the invention disclosed herein, internationalized software can be

developed and tested at much lower cost and with greater efficiency than was heretofore possible.

In this disclosure, there is shown and described only the preferred embodiment of the invention, but, as
5 aforementioned, it is to be understood that the invention is capable of use in various other combinations and environments and is capable of changes or modifications within the scope of the inventive concept as expressed herein.

FOOTNOTES